

## SEQUENCE LISTING

<110> Merck & Co., Inc.  
Jansen, Kathrin U.  
Schultz, Loren D.  
Neeper, Michael P.  
Markus, Henry Z.

<120> OPTIMIZED EXPRESSION OF HPV 31 L1 IN  
YEAST

<130> 21188-PCT

<150> 60/457,172

<151> 2003-03-24

<160> 8

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1515

<212> DNA

<213> HPV31 L1 wild-type

<400> 1

```

atgtctctgt ggcggcctag cgaggctact gtctacttac cacctgtccc agtgtctaaa 60
gttgtaagca cggatgaata tgtaacacga accaacatat attatcacgc aggcagtgc 120
aggctgctta cagtaggcca tccatattat tccataccta aatctgacaa tcctaaaaaa 180
atagttgtac caaagggtgc aggattacaa tatagggtat ttagggttcg ttaccagat 240
ccaaacaaat ttggatttcc tgatacatct ttttataatc ctgaaactca acgcttagtt 300
tgggcctgtg ttggtttaga ggtaggtcgc gggcagccat taggtgtagg tattagtggt 360
catccattat taaataaatt tgatgacact gaaaactcta atagatatgc cggtggtcct 420
ggcactgata ataggggaatg tataatcaatg gattataaac aaacacaact gtgtttactt 480
ggttgcaaac cacctattgg agagcattgg ggtaaaggta gtccttgtag taacaatgct 540
attaccctcg gtgattgtcc tccattagaa ttaaaaaaatt cagttataca agatggggat 600
atggttgata caggcttttg agctatggat tttactgctt tacaagacac taaaagtaat 660
gttccttttg acatttgtaa ttctatttgt aaatatccag attatcttaa aatgggtgct 720
gagccatatg gcgatacatt atttttttat ttacgtaggg aacaaatggt tgtaaggcat 780
ttttttaata gatcaggcac ggttggtgaa tcggtcctta ctgacttata tattaaggc 840
tccggttcaa cagctacttt agctaacagt acatactttc ctacacctag cggctccatg 900
gttacttcag atgcacaaat ttttaataaa ccataattgga tgcaacgtgc tcagggacac 960
aataatggta tttggtgggg caatcagtta tttgttactg tggtagatac cacacgtagt 1020
accaatatgt ctgtttggtc tgcaattgca aacagtgata ctacatttaa aagtagtaat 1080
tttaaagagt atttaagaca tggtaggaa tttgatttac aatttatatt tcagttatgc 1140
aaaataacat tatctgcaga cataatgaca tatattcaca gtatgaatcc tgctattttg 1200
gaagattgga attttggtt gaccacacct ccctcagggt ctttgaggga tacctatagg 1260
tttgtaacct cacaggccat tacatgtcaa aaaagtgcc cccaaaagcc caaggaagat 1320
ccattttaaag attatgtatt ttgggaggtt aattttaaag aaaagttttc tgcagattta 1380
gatcagtttc cactgggtcg caaattttta ttacaggcag gatatagggc acgtcctaaa 1440
tttaaagcag gtaaacgtag tgcaccctca gcatctacca ctacaccagc aaaacgtaaa 1500
aaaactaaaa agtaa 1515

```

<210> 2

<211> 1515

<212> DNA

<213> Artificial Sequence

<220>

<223> 31 partial rebuild

&lt;400&gt; 2

atgtctctgt	ggcggcctag	cgaggctact	gtctacttac	cacctgtccc	agtgtctaaa	60
gttgtaagca	cggatgaata	tgtaacacga	accaacatat	attatcacgc	aggcagtgc	120
aggctgctta	cagtaggcca	tccatattat	tccataccta	aatctgacaa	tcctaaaaaa	180
atagttgtac	caaagggtgc	aggattacaa	tatagggtat	ttaggggttcg	tttaccagat	240
ccaaacaaat	ttggatttcc	tgatacatct	ttttataatc	ctgaaaactca	acgcttagtt	300
tgggcctgtg	ttgggtttaga	ggtaggtcgc	gggcagccat	taggtgtagg	tattagtgg	360
catccattat	taaataaatt	tgatgacact	gaaaactcta	atagatatgc	cgggtggctc	420
ggcactgata	ataggggaatg	tatatcaatg	gattataaac	aaacacaact	gtgtttactt	480
gggtgcaaac	cacctattgg	agagcattgg	ggtaaaggta	gtcctttag	taacaatgct	540
attaccocctg	gtgattgtcc	tccattagaa	ttaaaaaatt	cagttataca	agatggggat	600
atgggttgata	caggcttttg	agctatggat	tttactgctt	tacaagacac	taaaagtaat	660
gttccttttg	acatttgtaa	ttctatttgg	aaatatccag	attatcttaa	aatgggtgct	720
gagccatacg	gcgacacctt	gttcttctat	ttgcgtagag	aacagatggt	cgtaaggcac	780
ttcttcaaca	gatccggcac	cgtagggtgaa	tctgtcccaa	ccgacctgta	catcaagggc	840
tccgggtcca	ccgctaccct	ggctaactcc	acctacttcc	caactccatc	tggctccatg	900
gtcacctccg	acgctcagat	cttcaacaag	ccatactgga	tgcagcgtgc	acagggtcac	960
aacaacggta	tctgttgggg	taaccagctg	ttcgtgactg	tggtcgatac	cacgcgttct	1020
accaacatgt	ctgtctgtgc	tgcaatcgct	aactctgaca	ctaccttcaa	gtcctctaac	1080
ttcaaggagt	acctgagaca	tggtgaggaa	ttcgatctgc	aattcatctt	ccagttgtgc	1140
aagatcaccc	tgtctgctga	catcatgacc	tacatccaca	gtatgaaccc	tgccatcctg	1200
gaggactgga	acttcgggtc	gaccactcca	ccttcgggtt	ctttggagga	tacctatagg	1260
tttgtaacct	cacaggccat	tacatgtcaa	aaaagtgcct	cccaaaagcc	caaggaagat	1320
ccatttaaag	attatgtatt	ttgggaggtt	aattttaaag	aaaagtttct	tgcagattta	1380
gatcagtttc	cactgggtcg	caaattttta	ttacaggcag	gatatagggc	acgtcctaaa	1440
tttaaagcag	gtaaacgtag	tgcaccctca	gcactctacca	ctacaccagc	aaaacgtaaa	1500
aaaactaaaa	agtaa					1515

&lt;210&gt; 3

&lt;211&gt; 1515

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; 31 total rebuild

&lt;400&gt; 3

atgtctttgt	ggagaccatc	tgaagctacc	gtctacttgc	caccagtccc	agtctctaag	60
gtcgtctcta	ccgacgaata	cgtcaccaga	accaacatct	actaccacgc	tgggtctgct	120
agattgttga	ccgtcgggtca	cccatactac	tctatcccaa	agtctgacaa	cccaaagaag	180
atcgtcgtcc	caaagggtctc	tggtttgcaa	tacagagtct	tcagagttag	attgccagac	240
ccaaacaagt	tcggtttccc	agacacctct	ttctacaacc	cagaaaccca	aagattgggtc	300
tgggcttgtg	tcggttttgga	agtcggtaga	ggccaaccat	tgggtgtcgg	tatctctggt	360
caccatttgt	tgaacaagtt	cgacgacacc	gaaaactcta	acagatacgc	tgggtggcca	420
ggtaccgaca	acagagaatg	tatctctatg	gactacaagc	aaacccaatt	gtgtttgttg	480
ggttgtaagc	caccaatcgg	tgaacactgg	ggtaagggtt	ctccatgttc	taacaacgct	540
atcaccccag	gtgactgtcc	accattggaa	ttgaagaact	ctgtcatcca	agacgggtgac	600
atgggtcgaca	ccggtttccg	tgtatggac	ttcaccgctt	tgcaagacac	caagtctaac	660
gtcccatttg	acatctgtaa	ctctatctgt	aagtaccag	actacttgaa	gatggtcgct	720
gaaccatacg	gcgacacctt	gttcttctac	ttgcgtagag	aacagatggt	cgtaaggcac	780
ttcttcaaca	gatccggcac	cgtagggtgaa	tctgtcccaa	ccgacctgta	catcaagggc	840
tccggttcca	ccgctaccct	ggctaactcc	acctacttcc	caactccatc	tggctccatg	900
gtcacctccg	acgctcagat	cttcaacaag	ccatactgga	tgcagcgtgc	acagggtcac	960
aacaacggta	tctgttgggg	taaccagctg	ttcgtgactg	tggtcgatac	cacgcgttct	1020
accaacatgt	ctgtctgtgc	tgcaatcgct	aactctgaca	ctaccttcaa	gtcctctaac	1080
ttcaaggagt	acctgagaca	tggtgaggaa	ttcgatctgc	aattcatctt	ccagttgtgc	1140
aagatcaccc	tgtctgctga	catcatgacc	tacatccaca	gtatgaaccc	tgccatcctg	1200
gaggactgga	acttcgggtc	gaccactcca	ccttcgggtt	ctttggaaga	cacctacaga	1260
ttcgtcacct	ctcaagctat	cacctgtcaa	aagtctgctc	cacaaaagcc	aaaggaagac	1320
ccattcaagg	actacgtctt	ctgggaagct	aaagtctctc	aaaagttctc	tgtgacttg	1380
gaccaattcc	cattgggtag	aaagtctctg	ttgcaagctg	gttacagagc	tagaccaaa	1440
ttcaaggctg	gtaagagatc	tgtctcatct	gcttctacca	ccacccagc	taagagaaag	1500

aagaccaaga agtaa

1515

&lt;210&gt; 4

&lt;211&gt; 504

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; HPV 31 L1

&lt;400&gt; 4

```

Met Ser Leu Trp Arg Pro Ser Glu Ala Thr Val Tyr Leu Pro Pro Val
 1          5          10          15
Pro Val Ser Lys Val Val Ser Thr Asp Glu Tyr Val Thr Arg Thr Asn
          20          25          30
Ile Tyr Tyr His Ala Gly Ser Ala Arg Leu Leu Thr Val Gly His Pro
          35          40          45
Tyr Tyr Ser Ile Pro Lys Ser Asp Asn Pro Lys Lys Ile Val Val Pro
          50          55          60
Lys Val Ser Gly Leu Gln Tyr Arg Val Phe Arg Val Arg Leu Pro Asp
          65          70          75          80
Pro Asn Lys Phe Gly Phe Pro Asp Thr Ser Phe Tyr Asn Pro Glu Thr
          85          90          95
Gln Arg Leu Val Trp Ala Cys Val Gly Leu Glu Val Gly Arg Gly Gln
          100          105          110
Pro Leu Gly Val Gly Ile Ser Gly His Pro Leu Leu Asn Lys Phe Asp
          115          120          125
Asp Thr Glu Asn Ser Asn Arg Tyr Ala Gly Gly Pro Gly Thr Asp Asn
          130          135          140
Arg Glu Cys Ile Ser Met Asp Tyr Lys Gln Thr Gln Leu Cys Leu Leu
          145          150          155          160
Gly Cys Lys Pro Pro Ile Gly Glu His Trp Gly Lys Gly Ser Pro Cys
          165          170          175
Ser Asn Asn Ala Ile Thr Pro Gly Asp Cys Pro Pro Leu Glu Leu Lys
          180          185          190
Asn Ser Val Ile Gln Asp Gly Asp Met Val Asp Thr Gly Phe Gly Ala
          195          200          205
Met Asp Phe Thr Ala Leu Gln Asp Thr Lys Ser Asn Val Pro Leu Asp
          210          215          220
Ile Cys Asn Ser Ile Cys Lys Tyr Pro Asp Tyr Leu Lys Met Val Ala
          225          230          235          240
Glu Pro Tyr Gly Asp Thr Leu Phe Phe Tyr Leu Arg Arg Glu Gln Met
          245          250          255          260
Phe Val Arg His Phe Phe Asn Arg Ser Gly Thr Val Gly Glu Ser Val
          260          265          270
Pro Thr Asp Leu Tyr Ile Lys Gly Ser Gly Ser Thr Ala Thr Leu Ala
          275          280          285
Asn Ser Thr Tyr Phe Pro Thr Pro Ser Gly Ser Met Val Thr Ser Asp
          290          295          300
Ala Gln Ile Phe Asn Lys Pro Tyr Trp Met Gln Arg Ala Gln Gly His
          305          310          315          320
Asn Asn Gly Ile Cys Trp Gly Asn Gln Leu Phe Val Thr Val Val Asp
          325          330          335
Thr Thr Arg Ser Thr Asn Met Ser Val Cys Ala Ala Ile Ala Asn Ser
          340          345          350
Asp Thr Thr Phe Lys Ser Ser Asn Phe Lys Glu Tyr Leu Arg His Gly
          355          360          365
Glu Glu Phe Asp Leu Gln Phe Ile Phe Gln Leu Cys Lys Ile Thr Leu
          370          375          380
Ser Ala Asp Ile Met Thr Tyr Ile His Ser Met Asn Pro Ala Ile Leu
          385          390          395          400
Glu Asp Trp Asn Phe Gly Leu Thr Thr Pro Pro Ser Gly Ser Leu Glu

```

				405				410					415				
Asp	Thr	Tyr	Arg	Phe	Val	Thr	Ser	Gln	Ala	Ile	Thr	Cys	Gln	Lys	Ser		
			420					425					430				
Ala	Pro	Gln	Lys	Pro	Lys	Glu	Asp	Pro	Phe	Lys	Asp	Tyr	Val	Phe	Trp		
		435					440					445					
Glu	Val	Asn	Leu	Lys	Glu	Lys	Phe	Ser	Ala	Asp	Leu	Asp	Gln	Phe	Pro		
	450					455					460						
Leu	Gly	Arg	Lys	Phe	Leu	Leu	Gln	Ala	Gly	Tyr	Arg	Ala	Arg	Pro	Lys		
465					470					475					480		
Phe	Lys	Ala	Gly	Lys	Arg	Ser	Ala	Pro	Ser	Ala	Ser	Thr	Thr	Thr	Pro		
				485				490						495			
Ala	Lys	Arg	Lys	Lys	Thr	Lys	Lys										
			500														

<210> 5  
 <211> 34  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> PCR Primer

<400> 5  
 cgctcgacgta aacgtgtatc atattttttt acag 34

<210> 6  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> PCR Primer

<400> 6  
 cagacacatg tattacatac acaac 25

<210> 7  
 <211> 41  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> PCR Primer

<400> 7  
 ctcagatctc acaaaacaaa atgtctctgt ggcggcctag c 41

<210> 8  
 <211> 38  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> PCR Primer

<400> 8  
 gacagatctt acttttttagt ttttttacgt tttgctgg 38